Assignment 6
Theoretical Neuroscience

Maneesh Sahani, Peter Dayan

November 19, 2011

1. **Estimation Theory**

   (a) We derived the Fisher information $J(\theta)$ as the expected value of the second derivate (curvature) of the log-likeihood in the lecture.

      i. Repeat the derivation for a *vector* parameter (or stimulus in our setting) $\vec{\theta}$, showing that the Fisher information in this case is given by a matrix.

   As mentioned in the lecture, there is an alternate definition in terms of the first derivative. For vector parameters this is:

$$J(\vec{\theta}_0) = \mathrm{Cov}_{\vec{\theta}_0}\left(\nabla \log p(n|\vec{\theta}) \Big|_{\vec{\theta}_0} \right).$$

   where $\mathrm{Cov}_{\vec{\theta}_0}$ means the covariance evaluated under $p(n|\vec{\theta}_0)$.

      ii. Demonstrate that these two definitions are the same (or more precisely, give conditions under which these two definitions are the same).

   (b) Consider an LNP model:

$$p(n|\vec{x}) = \mathsf{Poiss}(g(\vec{w} \cdot \vec{x}))$$

      i. What is $J(\vec{x})$ (the Fisher Information about the stimulus value available to the rest of the brain)? How does it depend on $\vec{w}$? Working in two dimensions (recall the picture from lecture) show how $J(\vec{x})$ varies around the vector linear projection vector $\vec{w}$.

      ii. What is $J(\vec{w})$ (the Fisher Information about the weight vector available to an experimenter — consider the case of multiple measurements n_i, each in response to a different stimulus $\vec{x}_i$)? How does it depend on the distribution of $\vec{x}$? What would be a good distribution with which to probe the cell if we knew (say) the orthant of stimulus space in which $\vec{w}$ lay?

2. **Population Coding**

   Shadlen and collaborators have claimed that if the activities of neurons in population codes are corrupted by *correlated* noise, then there is a sharp limit to the useful number of neurons in the population. *Prima facie* this is wrong – the stronger the correlations, the lower the entropy of the noise, and therefore the stronger the signal.

   Resolve this issue for the case of additive and multiplicative noise by considering the following three models for the noisy activities r_1 and r_2 of two neurons which form a population code for a real-valued quantity x:

$$\text{a)} \quad \begin{cases} r_1^a = x + \epsilon_1 \\ r_2^a = x + \epsilon_2 \end{cases} \tag{1}$$

$$\text{b)} \quad \begin{cases} r_1^b = x(1-\delta) + \epsilon_1 \\ r_2^b = x(1+\delta) + \epsilon_2 \end{cases} \tag{2}$$

$$\text{c)} \quad \begin{cases} r_1^c = x(1-\delta)(1+\eta_1) \\ r_2^c = x(1+\delta)(1+\eta_2) \end{cases} \tag{3}$$

   where $\delta \neq 0$ is known, and, ϵ and η are Gaussian, with mean 0 and covariance matrices:

$$\Sigma = \begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix}$$

(a) What is the maximum likelihood estimator (MLE) for x on the basis of r_1 and r_2 in each case?

(b) How does the expected accuracy in each case depend on the degree of correlation c? [Hint: begin by showing that the Fisher Information for a Gaussian distribution with mean $\mu(\theta)$ and variance $\Sigma(\theta)$ both dependent on a scalar parameter θ is:

$$J(\theta) = \nabla\mu^{\intercal}\Sigma^{-1}\nabla\mu + \frac{1}{2}\mathrm{Tr}\big[\Sigma^{-1}(\nabla\Sigma)\Sigma^{-1}(\nabla\Sigma)\big]$$

where the matrix "gradient" is the matrix of elementwise derivatives.]

(c) What conclusions would you draw about the clash between Shadlen and common sense?

3. **Temporal Difference learning**

(a) Simulate the operation of a temporal difference learning rule for the case of Figure 9.2A in the textbook. Discretize time at a resolution of 5ms for the temporal kernel of Eqn 9.6.

(b) Consider the case of partial reinforcement in which the stimulus comes on on every trial, but the reward only appears with probability p. What is the asymptotic variance of the prediction error at times between t=100 and 200ms? [An empirical result is enough for this question]

(c) If the dopaminergic code for TD prediction error is asymmetric:

$$\mathrm{DA}_t = \begin{cases} \delta_t & \text{if } \delta_t > 0 \\ \frac{1}{6}\delta_t & \text{if } \delta_t < 0 \end{cases}$$

What effect does the asymptotic variance in (b) have on the mean dopamine signal across the trial?